

**Univerzita Karlova v Praze**

**Filozofická fakulta**

Ústav informačních studií a knihovnictví

# **Bakalářská práce**

Marek Čejka

**Vyhledávací systémy nové generace**

New Generation Search Engines

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze dne .....

.....

Marek Čejka

## **Abstrakt**

Bakalářská práce představuje nový způsob vyhledávání a vyhledávací systémy nové generace, které toto vyhledávání zprostředkovávají.

V úvodní části je nový koncept vyhledávání, označovaný termínem „web scale discovery“, obecně představen a definován tak, že nový přístup tzv. web scale discovery systémů k vyhledávání spočívá v důrazu na rozsáhlý centrální index záznamů a okamžité rychlé vyhledávání spojující výsledky z mnoha elektronických databází a také lokálních zdrojů knihovny v rámci jediného jednoduchého vyhledávacího rozhraní. Je také vysvětleno, jakým způsobem jsou tyto centrální indexy tvořeny a jaké jsou nevýhody tohoto řešení.

Jádro práce tvoří nejprve jednotlivé představení a analýza vybraných komerčních web scale discovery systémů a následně jejich srovnání z hlediska funkčních možností. Představeny jsou následující systémy: EBSCO Discovery Service, Primo a Summon. Každý z nich je uveden stručnou charakteristikou, kterou následuje informace o obsahu centrálního indexu a samotné představení jednotlivých prvků a nabízených funkcí na příkladu vybrané fungující implementace. Představení každého systému uzavírá shrnutí jeho nejvýraznějších vlastností.

Závěrečná kapitola porovnává podporu vybraných funkcí u všech analyzovaných systémů v přehledných tabulkách s doplňujícím slovním komentářem.

V závěru jsou shrnuty nabyté poznatky včetně celkového zhodnocení užitečnosti a efektivity představených vyhledávacích systémů nové generace.

## **Klíčová slova**

vyhledávací systémy – web scale discovery – centrální index – knihovny – EBSCO Discovery Service – Primo – Summon

## **Abstract**

This thesis introduces a new method of information retrieval and search engines of a new generation which provide it.

The opening chapter introduces the new concept of searching and information retrieval, called “web scale discovery”, and defines the so called web scale discovery services as services that quickly search through a large central index of content compiled from both electronic databases and local library collections using a single simple search interface. It is also explained how central indexes are produced and what their main disadvantages are.

The core of the thesis is first dedicated to presenting and analysing individual selected commercial web scale discovery services and then comparing them in terms of functional capabilities. Presented are the following services: EBSCO Discovery Service, Primo and Summon. A chapter for each of the services begins with a brief overview, which is followed by a description of their central index and its content, and then presents individual features on an example of a working implementation of the service. Each chapter is concluded with a summary of the most distinctive features.

The last chapter compares support of selected functions in analyzed search engines using tables and supplementary commentary.

The conclusion summarises gained knowledge including an overall assessment of usefulness and effectiveness of the presented new generation search engines.

## **Keywords**

search engines – web scale discovery – central index – libraries – EBSCO Discovery Service – Primo – Summon

## Obsah

Předmluva .....	7
1 Vyhledávání nové generace.....	9
1.1 Motivace pro novou generaci vyhledávání .....	9
1.2 Web scale discovery – nový přístup k vyhledávání .....	9
1.3 Definice web scale discovery .....	10
1.3.1 Web scale .....	11
1.3.2 Web scale discovery .....	11
1.4 Obsah centrálních indexů .....	13
1.5 Úskalí konceptu web scale discovery a centrálních indexů .....	14
2 Analýza vybraných komerčních web scale discovery systémů.....	16
2.1 Úvod.....	16
2.2 EBSCO Discovery Service .....	16
2.2.1 Charakteristika systému .....	16
2.2.2 Obsah.....	17
2.2.3 Možnosti vyhledávání .....	17
2.2.4 Výsledky vyhledávání .....	20
2.2.5 Zobrazení záznamu.....	21
2.2.6 Další funkce.....	22
2.2.7 Shrnutí .....	22
2.3 Primo .....	23
2.3.1 Charakteristika systému .....	23
2.3.2 Obsah.....	24
2.3.3 Možnosti vyhledávání .....	24
2.3.4 Výsledky vyhledávání .....	26
2.3.5 Zobrazení záznamu.....	28
2.3.6 Další funkce.....	29
2.3.7 Shrnutí .....	30

2.4	Summon .....	30
2.4.1	Charakteristika systému .....	30
2.4.2	Obsah.....	31
2.4.3	Možnosti vyhledávání .....	32
2.4.4	Výsledky vyhledávání .....	33
2.4.5	Zobrazení záznamu.....	35
2.4.6	Další funkce.....	36
2.4.7	Shrnutí .....	36
3	Srovnání vybraných systémů z hlediska funkčních možností.....	38
3.1	Obecné funkce systému.....	38
3.2	Funkce vyhledávání .....	39
3.3	Záznamy .....	40
3.4	Možnosti nastavení systému knihovnou .....	41
3.5	Možnosti nastavení systému uživatelem.....	42
4	Závěr.....	43
	Seznam použité literatury.....	45
	Seznam obrázků a tabulek.....	48
	Seznam zkratk .....	49

## Předmluva

Bakalářská práce pojednává o novém konceptu vyhledávání a vyhledávačů určených pro knihovny. Nový přístup označovaný anglickým termínem „web scale discovery“ je založen na systémech, ve kterých uživatel v rámci jednoho jednoduchého rozhraní vyhledává v široké nabídce převážně odborných zdrojů z mezinárodních databází i lokálních zdrojů knihovny sdružených do jednoho velkého centrálního indexu.

Cílem práce je obecně představit celý nový koncept s jeho hlavními výhodami a nevýhodami, ukázat tři významné příklady z komerční sféry a srovnat je tak, aby bylo patrné, že navzdory jejich mnohdy stejné základní funkcionalitě se v některých detailech liší.

Práce si neklade za cíl vyčerpávajícím způsobem představit všechny web scale discovery systémy zcela dopodrobna anebo je srovnat až v takové míře detailu, která by mohla profesionálním způsobem pomoci knihovnám při jejich výběru. Stejně tak nesrovnává obsahy centrálních indexů jednotlivých systémů, neboť ve většině případů není kompletně veřejně dostupný a jejich objem se neustále zvětšuje. V rámci možností této práce také není zjistit, zdali web scale discovery systémy skutečně pomáhají uživatelům efektivněji vyhledávat, ačkoliv takový výzkum by byl v českém prostředí jistě také vítaný.

K výběru tématu mě vedl nejen všeobecný zájem o nové technologie na poli informační vědy a knihovnictví, ale také dřívější setkání s web scale discovery systémy testovanými na Univerzitě Karlově a jejich aktivní – a úspěšné – použití při vyhledávání zdrojů pro tvorbu seminárních prací.

Pro praktický popis a analýzu vybraných systémů jsem volil pro osobní testování ty implementace systémů, do kterých mám jako uživatel příslušné instituce plný přístup a mohl jsem tedy zkoušet i funkce dostupné až po přihlášení. Byť dva z testovaných systémů byly pouze v ukázkové zkušební verzi, hodnocení to nijak neovlivnilo, neboť nabízí téměř kompletní funkcionalitu, jen nejsou plně nastaveny „na míru“, jako kdyby byly již placené.

První část práce se zabývá obecným představením web scale discovery. Od motivace pro další pokrok na poli elektronického vyhledávání v knihovnách, přes vznik samotného termínu a jeho definování až po vysvětlení konceptu centrálních indexů a jejich možných nedostatků.

Jádro práce tvoří popis a analýza tří vybraných komerčních web scale discovery systémů na konkrétně zvolených případech implementace. U každého systému je představena jeho obecná charakteristika a obsah indexu a následuje popis jednotlivých funkčních oblastí od možností vyhledávání, přes zobrazení výsledků a záznamů po popis případných dalších funkcí a závěrečné shrnutí.

Závěrečná kapitola porovnává v přehledných tabulkách jednotlivé funkce v daných kategoriích s důrazem na jasné srovnání jejich podpory mezi jednotlivými systémy. V samotném závěru jsou kompletně shrnuty nabyté poznatky včetně celkového zhodnocení užitečnosti a efektivity představených vyhledávacích systémů nové generace.

Web scale discovery systémy se velkou rychlostí vyvíjejí a zlepšují, a proto placené implementace již mohly být v době předkládání bakalářské práce aktualizovány na novější verze a testované zkušební implementace již nemusí být dostupné. Popis funkcí a jejich porovnání bylo vytvářeno podle verzí, které byly dostupné minimálně do března 2013.

Chtěl bych poděkovat své vedoucí PhDr. Anně Stöcklové za cenné rady a připomínky při zpracovávání bakalářské práce a také českým zástupcům producentů jednotlivých systémů, se kterými jsem byl v průběhu tvorby práce v kontaktu a kteří mi poskytli cenné informace o svých produktech. Jmenovitě patří díky Pavlu Synkovi (EBSCO Publishing – EBSCO Discovery Service), Vladimíru Karenovi (Alberta i come Praha – Summon) a Martinu Vojnarovi (Multidata – Primo).

Oproti předběžné osnově stanovené v zadání práce nedošlo k žádným změnám ohledně cíle nebo náplně práce kromě pozměnění názvů některých kapitol.

Použitá literatura je citována podle normy ČSN ISO 690:2010. Citace v textu jsou uváděny podle Harvardského systému.



# **1 Vyhledávání nové generace**

## **1.1 Motivace pro novou generaci vyhledávání**

Knihovny často vynakládají velké prostředky, aby svým čtenářům nabídly co největší množství elektronických zdrojů, jejichž cena se často každým rokem zvyšuje. Na pozadí těchto nákupů zcela jistě stojí naděje, aby byly tyto zdroje v co největší míře využívány, protože proč by je jinak měla knihovna předplácet? Ovšem aby je uživatelé knihoven využívali, musí být schopni je objevit. A jak se ukazuje, v dnešní době čím dál více záleží také na tom, aby je objevili co nejrychleji a nejsnadněji (Vaughan, 2011a).

Uživatelé dnešní doby jsou z používání internetových vyhledávačů typu Google zvyklí na jednoduché, rychlé vyhledávání poskytující okamžité výsledky, a knihovna (nebo jí poskytované služby) pro ně často nepředstavuje první výchozí bod, kde začínají vyhledávat informace. Velice dobrý popis takové situace představuje následující citace z výzkumu o vylepšování uživatelského přístupu ke knihovním katalogům: „Lidé nejenže používají informace, které je snadné najít; oni dokonce používají informace, o kterých ví, že jsou málo kvalitní a méně hodnověrné – dokud to nevyžaduje přílišnou námahu je najít - spíše než by použili informace, o kterých ví, že jsou kvalitnější a hodnověrnější, ale je těžší je najít.“ (Bates, 2003, s. 4).

Jak již bylo řečeno, knihovny vyhrazují velkou část svých finančních zdrojů na předplacení odborných databází a dalších komerčních elektronických zdrojů. Často pak také věnují velkou pozornost statistikám prozrazujícím, jak jsou tyto zdroje využívány. Zdroje, o které je ze strany uživatelů zájem malý, je pak těžké obhájit při rozhodování, zdali jejich předplatné o rok prodloužit.

Proto by knihovny měly věnovat pozornost všem způsobům, jakým pro své uživatele zlepšit přístup k celé šíři zdrojů a především informacím, které obsahují. Protože „aby byl obsah využíván, musí být možné jej objevit – a pro mnoho dnešních uživatelů, snadné jej objevit“ (Vaughan, 2011a, s. 8).

## **1.2 Web scale discovery – nový přístup k vyhledávání**

Okamžité vyhledávání napříč všemi zdroji knihovny pomocí jednoho jediného okénka – toto a ještě něco navíc slibuje nový přístup k vyhledávání skrývajícím se pod anglickým novotvarem web scale discovery. Myšlenka existující již od vzniku federativního vyhledávání, ovšem nyní i s rychlostí a uživatelskou přívětivostí Googlu.

Inovativnost web scale discovery nespočívá v technologiích, na kterých jsou nové systémy postaveny, nýbrž na novém způsobu přístupu k celé podstatě vyhledávání. Namísto prohledávání jednotlivých roztroušených databází skrze jedno unifikované rozhraní stojí podstata těchto nových systémů na prohledávání jednoho ohromného indexu (rejstříku) obsahu sestaveného jednak z komerčních elektronických informačních zdrojů, na základě dohod tvůrců web scale discovery systémů s vydavateli a poskytovateli tohoto obsahu, dále co největšího množství volně dostupných elektronických zdrojů a nakonec lokálních zdrojů knihovny, jako jsou katalogy, digitální knihovny, repozitáře a podobně (Coufalová, 2009; Vaughan, 2012).

Možnost do různé míry indexovat obsah ze všech lokálních zdrojů a současně také moderní a přívětivější uživatelské rozhraní přinášejí už OPACy (veřejně přístupné online katalogy) nové generace od doby svého vzniku na začátku minulého desetiletí, ale právě indexování i vzdáleně uloženého obsahu a vytvoření jednotného centrálního indexu prakticky všech zdrojů, je to, čím se od nich novější systémy typu web scale discovery odlišují.

Pro web scale discovery systémy je typické, že jsou nabízeny metodou software jako služba („software as a service“) spíše než jako software pro lokální implementaci (Breeding, 2012a). Toto řešení mimo jiné umožňuje, že je možné rozsah poskytovaných služeb neustále rozšiřovat.

### **1.3 Definice web scale discovery**

S tím, jak více a více knihoven spouští v ostrých nebo testovacích verzích web scale discovery systémy hlavních komerčních poskytovatelů, je pravděpodobné, že se – především zahraniční - uživatelé knihoven již s některou z těchto služeb setkali. Vyhledávání zdrojů skrze jedno jediné vyhledávací okno pro ně není díky zkušenostem s vyhledávači typu Google žádná novinka – naopak je to něco, co již při kontaktu s knihovnou a jejími zdroji v dnešní době očekávají (Hoeppner, 2012). A tak se nejspíše ani nepozastaví nad tím, že jde v rámci knihoven o něco revolučního či přinejmenším evolučního. Samotný termín web scale discovery by však těmto uživatelům nejspíše nic neřekl.

Abychom přesněji pochopili podstatu těchto nových systémů, které přinášejí knihovnám zcela nový přístup k vyhledávání, je nutné si je přesněji vymezit. Terminologie web scale discovery v době vytváření této práce není nikterak pevně ukotvená, proto budou

v následujících řádcích představeny především definice jednotlivých autorů, kteří se fenoménem web scale discovery začali jako jedni z prvních seriózně zabývat.

Ve Spojených státech amerických Národní organizace pro tvorbu norem NISO (National Information Standards Organization) ustanovila pracovní skupinu Open Discovery Initiative, která má za úkol vytvořit normu NISO RP-19 typu „doporučené praxe“, jejíž cílem je „definovat nejlepší praktiky pro knihovní discovery systémy nové generace, které jsou založené na indexovaném vyhledávání“ (National Information Standards Organization, 2013). Projekt zahájený v srpnu roku 2011 má předpokládaný termín ukončení stanovený na srpen roku 2013, tedy až po dokončení této bakalářské práce. Proto bohužel nelze jeho výstupy v této práci využít.

### **1.3.1 Web scale**

Termín web scale (či web-scale) podle Marshalla Breedinga (2012a) poprvé zařadil do knihovnického slovníku Lorcan Dempsey, viceprezident a hlavním stratég americké neziskové společnosti OCLC (Online Computer Library Center, Inc.), který na něj v lednu 2007 upozornil v příspěvku na svém blogu. Poukázal na to, jak termín web scale již nějakou dobu využívají firmy typu Amazon ve spojení se svými novými internetovými výpočetními a úložnými službami. Termín web scale v takových případech odkazuje na to, že lze rozsah poskytovaných služeb rozšiřovat podle potřeby využití (Dempsey, 2007).

Společnost OCLC následně termín web scale začala používat v rámci svého konceptu globálního poskytování služeb a produktů, jako je katalog WorldCat, později discovery systému WorldCat Local a především služby Web-Scale Management Services (dnes již pod názvem OCLC Worldshare Management Services) – prvního knihovního systému pro kooperativní správu. OCLC nicméně na termín web scale nedrží žádná autorská práva, a proto se začal rychle šířit a začali jej přebírat i konkurenční společnosti, jako například Serials Solutions pro svůj discovery systém Summon.

### **1.3.2 Web scale discovery**

Slovní spojení web scale tak postupně dalo vzniknout celému termínu web scale discovery, který je ústředním tématem této práce. Přesné definice tohoto pojmu se různí, neboť jde o velice novou oblast, a tak různí autoři přicházejí pokaždé s lehce odlišnými definicemi. Všechny definice však spojují společné charakteristiky.

Jason Vaughan pro potřeby čísla časopisu Library Technology Reports, které je celé věnováno tematice web scale discovery, zmíněný termín definuje jako „službu schopnou rychle a souvisle vyhledávat ve velkém rozsahu předsklizeného<sup>1</sup> a indexovaného obsahu“ (2011a, s. 6). Sám ale pro potřeby svého pozdějšího článku používá lehce odlišnou definici, což dokazuje, že jde skutečně o novou a stále se rozvíjející oblast: „Služby typu web-scale discovery jsou flexibilní služby, které poskytují schopnost v rámci ohromného skladiště obsahu tento obsah rychle a souvisle objevovat, poskytovat a řadit dle relevance.“ (2012, s. 32).

Jiní autoři pak v rámci svých definicí kladou důraz na fakt, že web scale discovery představuje spojení velkého centrálního indexu obsahu a nástavby v podobě tzv. „discovery layer“ (česky volně přeloženo jako „vrstva pro objevování“). Samotný centrální index je definován jako: „Sbírka předsklizených a zpracovaných metadat a plných textů, které tvoří prohledávatelný obsah web scale discovery služby.“ (Hoeppner, 2012). Discovery layer označuje uživatelské rozhraní a především samotný vyhledávací systém, který prohledává a zobrazuje obsah získaný ze zdrojů knihovny, v případě web scale discovery systémů především z centrálního indexu. Athena Hoeppnerová používá pro web scale discovery následující souhrnnou definici: „Předsklizený centrální index doplněný o discovery layer s bohatými funkcemi, poskytující jednotné vyhledávání v lokálních, otevřených i předplácených zdrojích knihovny.“ (Hoeppner, 2012, s. 7).

Třetí varianta definice web scale discovery pak klade důraz právě na zmiňovaný centrální index. Česká terminologická databáze knihovnictví a informační vědy uvádí termín web scale discovery jako jeden z pěti anglických ekvivalentů termínu „centralizované vyhledávání“, který definuje následovně: „Typ sjednoceného vyhledávání, v jehož průběhu vyhledávač místo prohledávání jednotlivých zdrojů realizuje dotaz nad jednotným vyhledávacím indexovým souborem. Technologické řešení je založeno na předem vytvořeném centrálním indexu pro všechny prohledávané zdroje.“ (Národní knihovna České republiky, c2009).

Ve srovnání s výše zmíněnou problematikou definice se pak již jako pouze marginální problém jeví otázka, zdali se termín má psát „web-scale discovery“ či „web scale discovery“,

---

<sup>1</sup> „Předsklizené“ je českým překladem anglického originálu „pre-harvested“, který odkazuje na tzv. „harvesting“, metodu hromadného sběru dat, např. pro účel tvorby centrálního indexu.

neboť zahraniční autoři i samotní tvůrci systémů používají jednou tu, jednou onu variantu. Z významového hlediska tato jazyková nejednotnost nehraje žádnou roli.

V této bakalářské práci bude používán původní anglický termín „web scale discovery“ (ve tvaru bez pomlčky), pro označení samotných systémů (v zahraniční literatuře označovaných také jako *services* neboli služby) pak slovní spojení „web scale discovery systémy“ či pouze zkráceně „discovery systémy“ nebo jen „systémy“. Hlavním důvodem je zachování přesnosti označení a navázání na koncepci používanou zahraničními autory.

#### 1.4 Obsah centrálních indexů

Ohromné centrální indexy předsklizeného obsahu jsou z velké části tím, co z web scale discovery systémů dělá nový přístup k vyhledávání. Na rozdíl od federativního vyhledávání<sup>2</sup>, které na první pohled uživateli poskytuje stejnou službu, jsou výsledky v případě vyhledávání v centrálním indexu zobrazeny takřka okamžitě.

Obsah centrálního indexu se zpravidla skládá ze sklizeného obsahu lokálních zdrojů knihovny (záznamy z katalogu a repozitářů) a především z obsahu získaného na základě licenčních dohod uzavřených s vydavateli, kteří umožní přístup ke svým metadatům, eventuelně plným textům, za účelem indexace.

Klasické vyhledávače typu Google pracují na podobném principu, avšak jejich práce je jednodušší v tom, že indexují volně přístupný obsah internetu, zatímco tvůrci web scale discovery systémů musí pro své centrální indexy obsah získat licenčními dohodami s jednotlivými poskytovateli. Vzájemná rivalita existujících společností anebo nedůvěra vydavatelů k novému přístupu k vyhledávání pak vede k tomu, že prakticky žádný z centrálních indexů neobsahuje veškerý obsah. Typicky se děje to, že producent discovery systému, který je zároveň vlastníkem některé databáze, neposkytuje obsah této databáze svým konkurentům, a tak je dostupný pouze v centrálním indexu jeho vlastního systému (Hoeppner, 2012).

To je jedním z největších problémů celého konceptu. V současné době proto musí knihovny web scale discovery systém pečlivě vybírat nejen z hlediska nabízených funkcí, ale také tak, aby se obsah centrálního indexu co nejvíce kryl s obsahem, který si knihovna

---

<sup>2</sup> Koncept federativního vyhledávání je založen na systému, který po zadání dotazu uživatelem prohledává zvlášť jednotlivé vzájemně nezávislé elektronické zdroje a výsledky zobrazuje v jednotném uživatelském rozhraní. Celý proces je ovšem zdlouhavý, může trvat až několik desítek vteřin. Jedním z těchto systémů je např. Metalib společnosti Ex Libris (Krejčíř, 2011).

aktuálně jednotlivě předplácí a který považuje za důležitý pro své uživatele. Systém, který by neměl indexovaný obsah ze zrovna té v knihovně nejvíce využívané databáze, by nesplnil svůj účel, protože knihovna by sice měla předplacené plné texty, ale v daném discovery systému by jejich záznamy (a tedy i ony plné texty) nikdo nenašel.

Díky tomu, že se stejné dokumenty mohou nacházet ve více různých databázích, se ovšem může stát, že centrální index obsahuje záznam z jedné databáze a přitom plný text si knihovna předplácí z jiné (té, se kterou jinak tvůrce discovery systému nemá uzavřenou licenční dohodu) – díky linkovacímu nástroji, který na plný text do správné databáze odkáže, se pak vše propojí.

Najít ten správný systém, který by odpovídal obsahovým požadavkům, je pro knihovny skutečně složité, a je proto nutné si od producenta systému předem vyžádat důkladnou komparační obsahovou analýzu (Vaughan, 2011f).

## **1.5 Úskalí konceptu web scale discovery a centrálních indexů**

Přestože si knihovníci a odborníci využívající elektronické informační zdroje uvědomují, co se web scale discovery systémy snaží přinést, uvědomují si zároveň i jejich nedostatky a bojí se, že s nimi přijdou o některé funkce, na které jsou zvyklí ze své práce s tradičními systémy. Marshall Breeding (2012b) ve svém textu zmiňuje, že zatímco prvotní reakce běžných uživatelů s novými systémy bývají nadšené, mnoho knihovníků a odborných uživatelů je v některých případech vidí i jako krok zpět.

Přidávají se k nim i obavy poskytovatelů obsahu. Ti se bojí ztráty kontroly nad tím, v jaké výsledné podobě se jejich zdroje dostávají k uživatelům, a toho, aby nepoklesla jejich statistika využívání, což by mohlo vést k neprodloužení předplatného ze strany knihoven. Někteří vydavatelé a databázová centra raději nabízejí své zdroje pouze ve vlastním uživatelském rozhraní než v systémech jiných společností, kde nemohou nijak přímo ovlivnit pozici svého obsahu v rámci vyhledaných výsledků.

Linda Beebe, ředitelka pro databázi PsycINFO, významného zdroje z oboru psychologie, uvedla: „Pokud by PsycINFO bylo součástí discovery služeb, uživatelé by se mohli domnívat, že doopravdy prohledávají naši databázi, i když by se ve skutečnosti sotva dotkli jejího povrchu.“ (Kelley, 2012). Přesto je však databáze PsycINFO součástí systému EBSCO Discovery Service. K tomu Linda Beebe dodává: „EDS je jediný systém, který má integrovanou platformu pro doručování obsahu. Můžeme si být jisti, že pouze uživatelé, kteří

mají licencovaný přístup k našemu obsahu, vidí záznam z PsycINFO a máme zaručeno, že systém využití posiluje, nikoliv že je jenom nahrazuje.“ (Kelley, 2012). Pro přístup k záznamům databáze PsycINFO je nutné být nejen přihlášený do systému EDS, ale knihovna také musí databázi předplácet – jinak se v seznamu výsledků její záznamy nezobrazí.

Současná praxe zahraničních knihoven, které již některý z discovery systémů zavedly, bývá taková, že jej nabízí jako primární přístup k nabízeným zdrojům, ale stále ještě ponechávají v činnosti staré katalogy a tradiční systémy pro použití knihovníky a dalšími uživateli se speciálními potřebami. Zatímco do vývoje web scale discovery systémů a discovery rozhraní (resp. OPACů nové generace) je investováno mnoho prostředků a úsilí, staré katalogy (OPACy druhé generace) již žádná vylepšení nečekají. Jde tedy pouze o krátkodobé až střednědobé řešení (Breeding, 2012b).

## **2 Analýza vybraných komerčních web scale discovery systémů**

### **2.1 Úvod**

Cílem této kapitoly je obecně i prakticky představit tři ze současně nejvýznamnějších komerčních web scale discovery systémů – EBSCO Discovery Service, Primo (resp. Primo Central) a Summon. Jedná se o systémy, ze kterých Univerzita Karlova v době tvorby této bakalářské práce vybírá nejvhodnější systém pro zakoupení a následné použití v univerzitních knihovnách. Pro popis a testování funkčních možností bude využito testovacích verzí systémů zprovozněných pro potřeby výběru Univerzity Karlovy, s výjimkou systému Summon, jehož testovací verze v době tvorby práce nebyla kontinuálně dostupná – v jeho případě bude použita ostrá placená verze systému nasazená v únoru 2013 v Národní technické knihovně.

Ve sféře komerčních web scale discovery systémů existuje ze světového hlediska ještě jeden významný zástupce – systém WorldCat Local společnosti OCLC. Přestože jde též o komplexní web scale discovery řešení, nebude v této práci blíže analyzován. Důvodem je malá tradice produktů OCLC v České republice, neexistující české zastoupení, nemožnost systém otestovat z pohledu přihlášeného uživatele a také omezený rozsah bakalářské práce nedovolující tento - především v USA rozšířený - web scale discovery systém blíže představit.

### **2.2 EBSCO Discovery Service**

#### **2.2.1 Charakteristika systému**

Vývoj EBSCO Discovery Service (EDS) společnosti EBSCO Publishing začal v roce 2008. Od veřejného oznámení na jaře 2009 uběhl téměř rok a na začátku roku 2010 byl systém oficiálně uveden na trh. Je nabízen výhradně jako vzdáleně hostovaná služba, lokální implementace není možná.

V současné době si ve světě předplácí EBSCO Discovery Service více než 2100 institucí. V České republice jsou provozovány čtyři placené instalace. Typ předplatitelů EDS se liší podle jednotlivé oblasti či státu. Obecně řečeno jde především o akademické knihovny, korporace a případně vládní instituce (např. národní banky).

Zprovoznění systému od finálního podepsání objednávky trvá zpravidla do 4 týdnů, pokud knihovna předem dodá všechny potřebné podklady. Pokud je implementován i katalog instituce, trvá doba zprovoznění 4-6 týdnů. Cena systémů se odvíjí především od velikosti instituce (počtu uživatelů) a také toho, zda bude implementován i katalog (a na množství



obsahu ke zpracování). Aktualizace přicházejí cca 4x ročně v podobě větších updatů, menší aktualizace probíhají průběžně dle potřeby.

V České republice distribuuje produkt lokální zastoupení společnosti EBSCO, které také poskytuje technickou podporu při implementaci i následně v době provozu systému<sup>3</sup>.

### **2.2.2 Obsah**

Discovery systém EBSCO Discovery Service<sup>4</sup> využívá rozsáhlý centrální index a může též integrovat obsah z lokálních zdrojů knihovny (katalogu, repozitářů apod.), získaný metodami pro sklizení obsahu, jako je OAI-PMH nebo FTP, a zpracovat data schémat typu MARC, Dublin Core, XML nebo EAD. Získaná metadata se převedou do standardizovaného formátu EBSCOhost (Vaughan, 2011d).

Celkový počet dokumentů v centrálním indexu EDS není přesně znám. Nabízí však obsah od více než 20 tisíc vydavatelů a 70 tisíc vydavatelů monografií. Jedná se o cca 50 tisíc titulů periodik, 6 milionů knih, a téměř 1 milion CD/DVD. Dále je zahrnuto také cca 100 milionů novinových článků a 400 tisíc konferenčních sborníků. Další statisíce záznamů pak pocházejí z dalších zdrojů (šedá literatura, dizertace, apod.).

Díky tomu, že je společnost EBSCO sama agregátorem mnoha významných databází, propaguje jako svou velkou přednost, že poskytuje u vybraných zdrojů bohatá a detailní metadata včetně autorských abstraktů, klíčových slov anebo řízených předmětových hesel.

### **2.2.3 Možnosti vyhledávání**

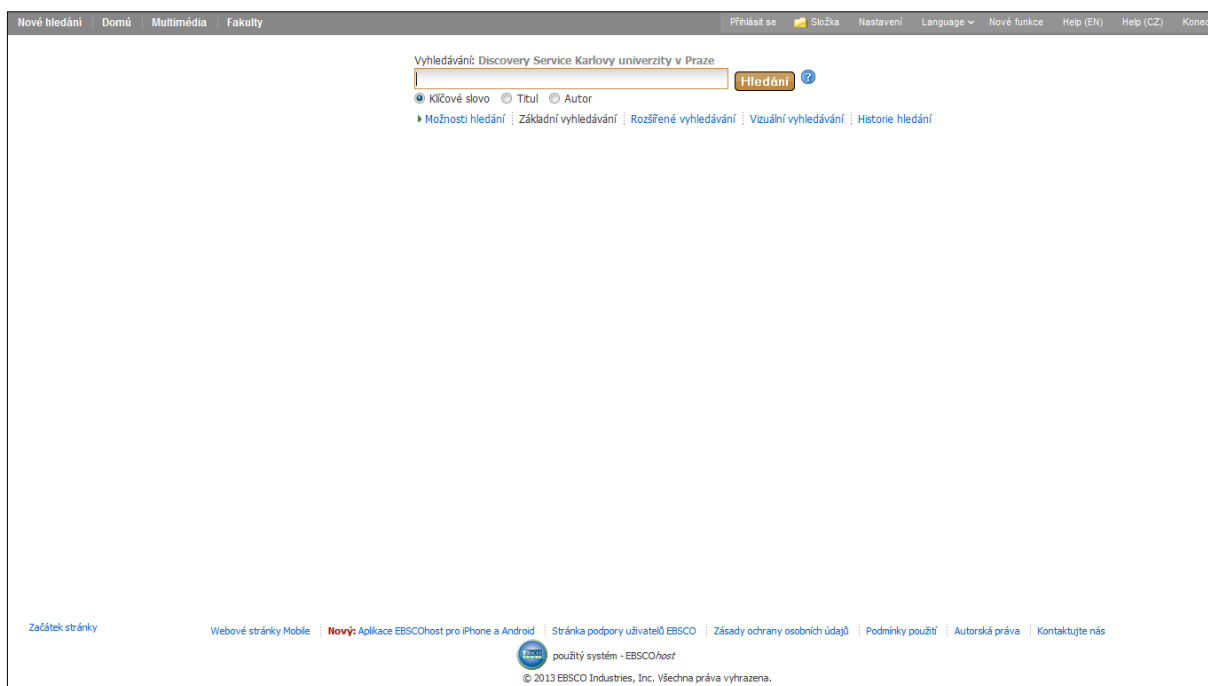
Použití web scale discovery systému EBSCO Discovery Service bude představeno na zkušební implementaci Univerzity Karlovy<sup>5</sup> (viz obr. 1).

---

<sup>3</sup> Necitované informace v kapitolách *Charakteristika systému* a *Obsah* poskytl Mgr. Pavel Synek, zástupce EBSCO Publishing, v emailové komunikaci.

<sup>4</sup> Oficiální webová stránka - <http://www.ebscohost.com/discovery>

<sup>5</sup> Nelze poskytnout přímý odkaz – pro přístup do testovací verze EDS na UK je nutné přihlášení nebo přístup z IP adres knihoven univerzity



Obr. 1: Úvodní stránka systému EDS

Přístup do systému EBSCO Discovery Service standardně vede skrze lokální univerzitní přístup (vázaný na specifické IP adresy) anebo přihlašovací stránku pro vzdálený přístup. V některých případech je u již spuštěných implementací dostupný tzv. „přístup pro návštěvníky“ (tzv. „Guest mode“), který poskytuje pouze omezené možnosti vyhledávání, především kvůli omezením týkajících se zobrazovaných výsledků (Vaughan, 2011d).

Uživatelské rozhraní sestává z vrchního ovládacího panelu, hlavního prostoru stránky určeného pro vyhledávání a doplňkových odkazů vespod. Nejdůležitější funkce horního ovládacího panelu se nacházejí v pravé části: přihlášení<sup>6</sup>, složka s uloženými záznamy, nastavení, jazyk a nápověda. Odkazy ve spodní části stránky jsou doplňkové: odkazují na mobilní verzi stránek, mobilní aplikaci anebo například podmínky použití.

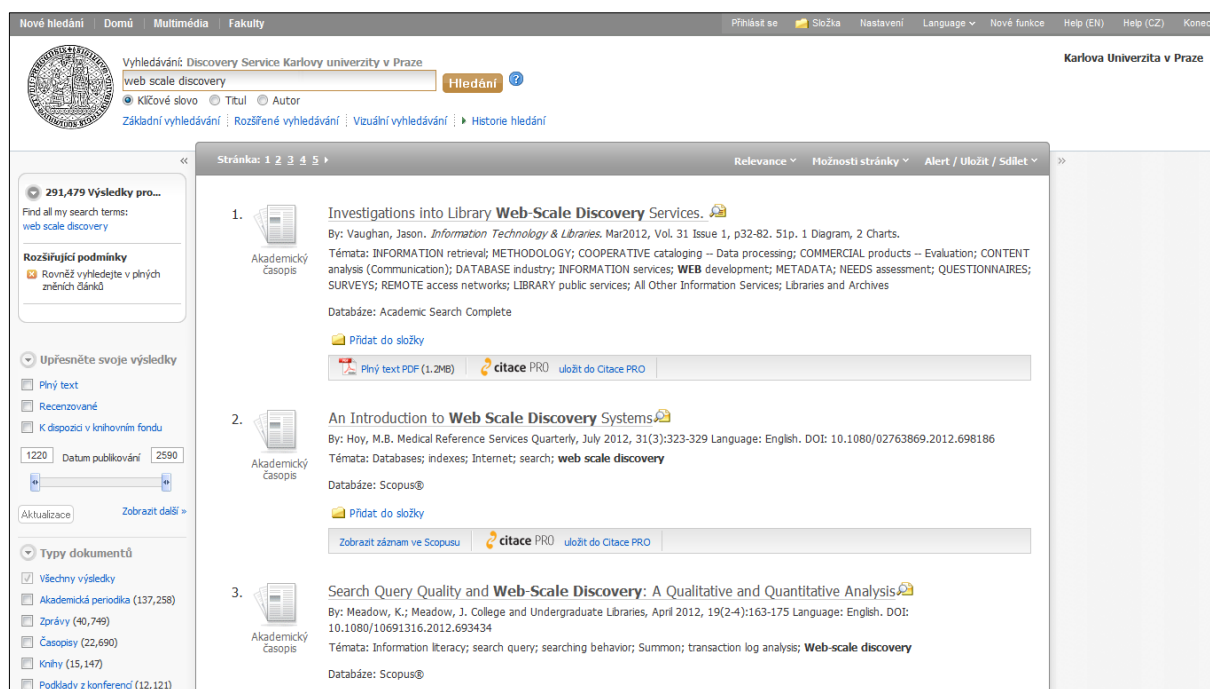
Primárním vyhledávacím rozhraním je v případě EDS jediné textové pole s aktivním tlačítkem nazvaným v české verzi *Hledání* (viz obr. 2). U něj se nachází ještě možnost pro omezení vyhledávání jen na pole autor a titul místo všeobecného vyhledávání na základě klíčových slov.

<sup>6</sup> Jedná se o přihlášení do účtu My EBSCOhost, který vyžaduje zvláštní bezplatnou registraci. Nejde o přihlášení pomocí knihovního/institucionálního účtu. Výhody účtu My EBSCOhost jsou popsány dále v práci.



## 2.2.4 Výsledky vyhledávání

Výsledky vyhledávání jsou v discovery systému EDS řazeny primárně podle relevance, alternativně podle data (od nejnovějších nebo od nejstarších). Největší část obrazovky zabírá seznam vyhledaných záznamů, po levé straně se nachází fasetová navigace a po pravé straně prázdný panel, určený mj. pro zobrazení dodatečných výsledků z volitelného federativního vyhledávání EBSCOhost Integrated Search (viz obr. 3).



Obr. 3: Ukázka výsledků vyhledávání v systému EDS

Vpravo nad seznamem vyhledaných výsledků se nachází nastavení *Možnosti stránky*, v rámci něhož si lze uživatelské rozhraní do značné míry upravit. Lze zvolit *Formát výsledku*, tedy obsáhlost informací zobrazovaných při vyhledávání v záznamu (jako výchozí je nastaven *Stručný záznam*), dále *Náhled obrazového materiálu v publikaci*, *Počet výsledků na stránku* a také *Rozvržení stránky*, v němž lze nastavit skrytí postranních panelů (fasetové navigace vlevo a pomocného panelu vpravo).

Vedle se nachází ještě aktivní odkaz *Alert / Uložit / Sdílet*, který zobrazí menu umožňující hromadné uložení výsledků nebo celého vyhledávacího dotazu do složky, vytvoření alertového upozornění na email nebo jako kanálu RSS, trvalý odkaz a dále ikonky pro sdílení na nespočet sociálních sítí.

Fasetová navigace po levé straně umožňuje uživateli vyfiltrovat výsledky vyhledávání pomocí řady kritérií. Je možné výsledky omezit pouze na ty s plným textem, recenzované

nebo ty, které jsou k dispozici ve fondu knihovny. Dále je možné omezení z časového hlediska, po rozkliknutí rozšířeného rozhraní až s přesností na měsíce. Další kritéria zahrnují omezení dle: typu dokumentu, tématu (předmětových hesel), vydavatele, publikace, jazyka, zeměpisu a zdrojů obsahu.

### **2.2.5 Zobrazení záznamu**

Standardně se vyhledané záznamy v EBSCO Discovery Service (stejně jako v klasických databázích EBSCO) zobrazují v tzv. stručném formátu. Pro každý typ obsahu se vedle záznamu zobrazuje odlišná obrazová ikona. U knih se, v případě dostupnosti, zobrazí místo standardní ikony obálka dané knihy<sup>10</sup>.

V případě knih obsahuje záznam typicky následující druh údajů: název, autor, informace o publikaci, rozsah stran, celkový počet stran, velikost a jazyk. U lokálních zdrojů se také zobrazuje informace o aktuální dostupnosti a lokaci dokumentu. Aktualizace tohoto údaje se provádí až dodatečně po zobrazení výsledků vyhledávání na uživatelsv dotaz a trvá několik vteřin navíc.

U článků závisí míra zobrazených údajů v záznamu na metadatech získaných ze zdrojové databáze. Standardními údaji jsou název, autor, publikace, ročník a číslo, rozsah stran a jejich počet. U mnoha článků jsou předmětová hesla, někde též prvních pár řádek abstraktu anebo náhledy obrázků a tabulek.

U všech druhů obsahu se v záznamu zobrazuje také informace o databázi (zdroji). U lokálních dokumentů bude uveden jako zdroj katalog, u článků to je konkrétní databáze. Nechybí odkaz na přidání záznamu do složky pro pozdější práci.

Po kliknutí na název je uživatel přenesen na novou stránku s podrobným záznamem a lehce odlišným rozhraním. Největší část obrazovky zabírá samotný záznam, ve kterém se nacházejí některé další dodatečné údaje jako je především abstrakt nebo například afiliace autora.

Po levé straně se nachází panel, skrze nějž lze přistupovat k plným textům, záznamům přímo v databázích poskytovatelů anebo konkrétně v testované verzi Univerzity Karlovy exportovat záznam do českého citačního generátoru Citace PRO.

---

<sup>10</sup> Všichni zákazníci EDS automaticky využívají obrazové databáze obálek společnosti Baker & Taylor. Pokud si knihovna od této společnosti (nebo od jiné) předplácí další obohacující zdroje, je možné je do EDS také připojit.

Jelikož je společnost EBSCO poskytovatelem řady významným databází a pokud si je instituce předplácí, lze zobrazit vybrané elektronické dokumenty přímo v nativním rozhraní systému EDS. Buď ve formě HTML textu anebo v PDF. Pokud je tato možnost dostupná, odkazují na ní příslušné odkazy ve zmiňovaném levém panelu.

Pravý panel nástrojů je věnovaný všem doplňkovým funkcím typu: přidat do složky, tisk, email, uložit, citovat, exportovat, vytvořit poznámku, trvalý odkaz a uložit do záložek.

#### **2.2.6 Další funkce**

Stejně jako v ostatních databázích EBSCO, i v rámci EBSCO Discovery Service lze využívat uživatelský profil My EBSCOhost. Ten je nezávislý na účtu uživatele v rámci jeho instituce a je pro něj nutná nová (bezplatná) registrace. Umožňuje využívat některé dodatečné funkce systému, jako například dlouhodobě ukládat vybrané záznamy, třídit se je do složek a dále s nimi pracovat (ukládat, exportovat do citačních manažerů atd.). Dále také upravit si natrvalo uživatelské rozhraní, jak již bylo představeno výše.

#### **2.2.7 Shrnutí**

Web scale discovery systém EBSCO Discovery Service vychází z konceptu existujících databází společnosti EBSCO. Nabízí podobný vzhled i funkce a na první pohled patří k těm složitějším discovery systémům. Tím ale může potěšit pokročilejší uživatele nebo knihovníky, kterým přináší jak všechny staré známé pokročilé funkce, tak všechny výhody velkého centrálního indexu obsahu. Záznamy často obsahují detailní metadata.

Nevýhodou je nutnost institucionálního přihlášení před samotným vyhledáváním, protože volný přístup v roli tzv. návštěvníka je velice omezený a pro reálné vyhledávání nepoužitelný. Jak ale bylo uvedeno dříve v práci, pro některé databáze (např. PsycINFO) je to naopak důvod, proč spolupracovat s EDS a s jinými producenty web scale discovery systémů.

Za související nevýhodu pak lze považovat fakt, že pro využívání pokročilých funkcí typu ukládání záznamů a pozdější práci s nimi nebo tvorbu alertových upozornění na email je nutné si vytvořit účet My EBSCOhost a nelze pro tyto účely použít institucionální účet.

## 2.3 Primo

### 2.3.1 Charakteristika systému

Na tvorbě systému Primo<sup>11</sup> začala společnost Ex Libris pracovat v roce 2005 a na trh byl oficiálně uveden v roce 2007. Systém Primo ale sám o sobě nenese všechny prvky web scale discovery, jde o tzv. discovery layer (viz kapitola Definice web scale discovery) a také software spadající do kategorie tzv. OPACů nové generace. Primo sklízí a indexuje lokální sbírky knihovny, jako jsou bibliografické záznamy, materiály z digitálních sbírek a repozitářů a nabízí jednotné vyhledávací rozhraní pro všechny tyto zdroje. Lze jej také nastavit pro vyhledávání v indexech vzdálených sbírek a spojit výsledky vyhledávání dohromady s lokálními (Vaughan, 2011e). Sám o sobě však neobsahuje centrální index nutný pro splnění definice web scale discovery.

Skutečnou službu typu web scale discovery přinesla společnost Ex Libris na trh v červnu 2010 uvedením produktu Primo Central<sup>12</sup>, který s sebou přináší globální centrální index. Spojením systému Primo a centrálního indexu Primo Central vznikne kompletní web scale discovery řešení od společnosti Ex Libris. Zatímco systém Primo lze provozovat lokálně i vzdáleně, Primo Central je poskytováno jako služba hostovaná výhradně vzdáleně na serverech firmy Ex Libris.

Nutno podotknout, že knihovny mohou provozovat pouze systém Primo jako OPAC nové generace anebo naopak zakoupit index Primo Central a k vyhledávání použít discovery vrstvu od jiné firmy. V rámci této práce bude Primo a Primo Central považováno za propojený celek, protože standardně je dodáváný objednavateli jako jeden komplexní web scale discovery systém, a budou představeny oba produkty dohromady.

V současné době využívá služeb systému Primo (tedy ne nutně kompletního web scale discovery řešení) více než 1900 institucí po celém světě (Ex Libris Ltd, 2012). V České republice je provozována jedna placená instalace web scale discovery systému Primo s indexem Primo Central. Hlavními předplatiteli jsou především knihovny velké (s velkým knihovním fondem a velkým počtem uživatelů), z hlediska typu především knihovny akademické, dále národní a odborné (special libraries).

---

<sup>11</sup> Oficiální webová stránka - <http://www.exlibrisgroup.com/category/PrimoOverview>

<sup>12</sup> <http://www.exlibrisgroup.com/category/PrimoCentral>

Zprovoznění Prima od finálního podepsání objednávky trvá od čtyř týdnů (v případě kompletní správy systému společností Ex Libris) po tři měsíce (v případě, že si knihovna chce systém spravovat sama). Nové verze systému přichází zhruba jednou za rok, rok a půl, dílčí aktualizace zhruba 4x za rok. Rychlé opravy jsou aplikovány podle potřeby.

V České republice distribuuje Primo společnost Multidata Praha s.r.o., která také poskytuje knihovnám technickou podporu při implementaci i následně v době provozu systému<sup>13</sup>.

### **2.3.2 Obsah**

Discovery systém Primo nabízí zaprvé obsah ze svého velkého centrálního indexu Primo Central a zadruhé možnost kompletní integrace lokálních zdrojů knihovny, ať už jde o katalog, digitální sbírky nebo repozitáře.

Zatímco index Primo Central je uložen vždy vzdáleně na serverech společnosti Ex Libris, u lokálních zdrojů závisí na variantě implementace, kterou knihovna zvolí. Buď mohou být data též nahrána do výpočetního oblaku (tzv. „cloudové řešení“) a být uložena vzdáleně, nebo mohou zůstat lokálně pod správou knihovny.

Primo dokáže získat obsah ze všech běžných knihovních katalogů a repozitářů. Podporuje různé metody sklizení lokálního obsahu včetně OAI-PMH a FTP, přičemž dokáže zpracovat data různých schémat: MARC/MARC XML, Dublin Core, EAD nebo MODS. Získaný obsah je zpracován do standardizovaného formátu Primo.

Centrální index Primo Central nabízí zhruba 500 milionů dokumentů, které pocházejí výhradně z odborných zdrojů. Společnost Ex Libris má uzavřené licenční dohody s mnoha významnými vydavateli, agregátory i open-access repozitáři.

Poté, co uživatel zadá vyhledávací dotaz, systém prohledá centrální index i lokální sbírky, výsledky skombinuje a zobrazí dohromady.

### **2.3.3 Možnosti vyhledávání**

Discovery systém Primo bude představen na zkušební implementaci Univerzity Karlovy<sup>14</sup> (viz obr. 4). Na začátek je nutné podotknout, že systém Primo je možné z hlediska vzhledu i nabízených funkcí v rámci uživatelského rozhraní do značné míry přizpůsobit,

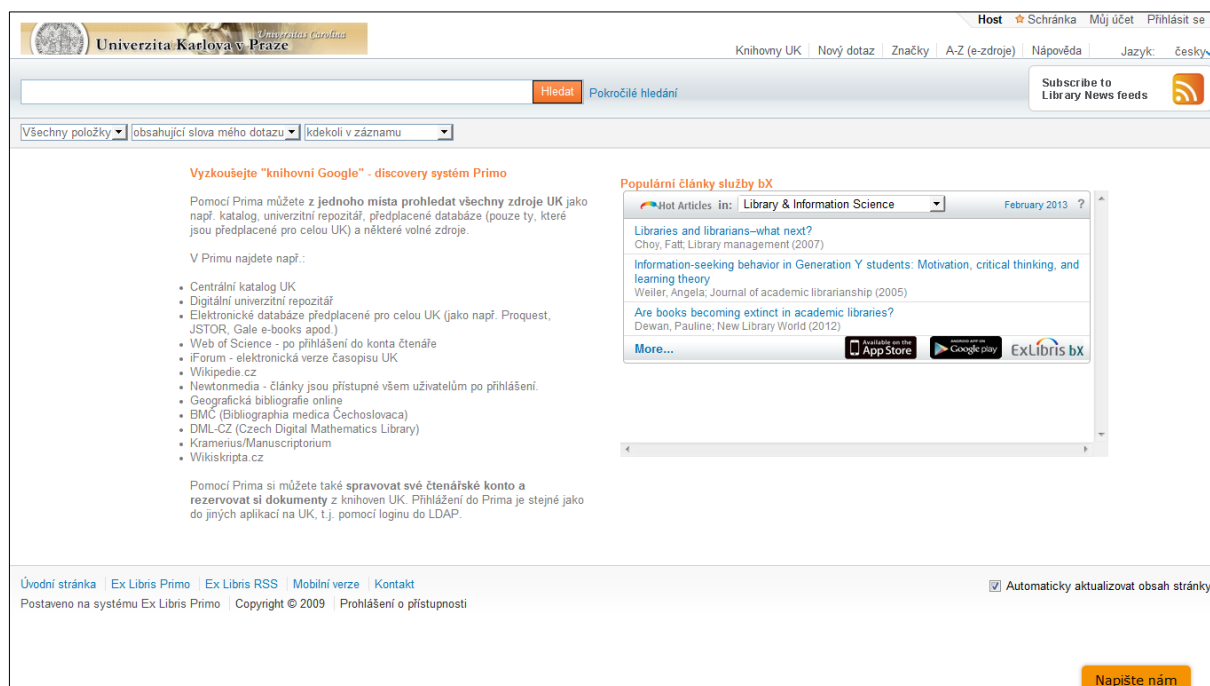
---

<sup>13</sup> Necitované informace v kapitolách *Charakteristika systému* a *Obsah* poskytl Martin Vojnar, vedoucí analytik ve firmě Multidata, při osobní a emailové komunikaci.

<sup>14</sup> <http://pelican.is.cuni.cz/>



neboť jde o výrazně flexibilní systém. Testovací verze UK využívá základní šablonu systému, a ta bude v následujících řádcích blíže popsána.



Obr. 4: Úvodní stránka systému Primo

Pro přístup do systému Primo není vyžadováno žádné přihlášení. Požadavky na autentizaci uživatele nastaví knihovna dle svých potřeb, standardně je přihlášení vyžadováno např. až v momentu pokusu o získání plného textu dokumentu anebo rezervace tištěného dokumentu (Vaughan, 2011e).

Uživatelské rozhraní discovery systému Primo obsahuje kromě vyhledávacího pole (popsáno dále v textu) hlavní uživatelský panel umístěný na všech stránkách v pravé horní části obrazovky a doplňkové panely v druhém řádku z vrchu a vespod. V rámci uživatelského panelu se zobrazuje identita uživatele (pokud není přihlášený, je uveden jako Host), lze skrze něj vstoupit do *Schránky* s uloženými záznamy, spravovat své konto pod záložkou *Můj účet* a přihlásit se. Pod touto nejsvrchnější lištou se nachází doplňkový ovládací panel, na kterém lze nastavit především jazyk systému, vstoupit do nápovědy a knihovna jej může využít pro odkazy dle svých specifických potřeb – v testované verzi obsahuje např. odkaz na *Knihovny UK*. Spodní panel poskytuje doplňkové odkazy, např. na mobilní verzi stránek.

K vyhledávání slouží primárně jednoduché základní vyhledávací rozhraní skládající se z jediného textového pole, aktivního tlačítka *Hledat* a odkazu na pokročilou verzi vyhledávacího rozhraní (viz obr. 5). Slouží k zadání dotazu na základě klíčových slov,

s možností omezení přesné fráze pomocí uvozovek a použití logických booleovských operátorů: AND, OR a NOT (zapsaných velkými písmeny). Lze také využít zástupných znaků „?“ a „\*“ pro nahrazení jednoho, respektive více znaků ve vyhledávaném termínu.

Obr. 5: Základní vyhledávací rozhraní systému Primo

V testovací verzi UK se pod základním rozhraním nachází ještě tři rolovací nabídky umožňující již dopředu dotaz zpřesnit na určitý typ dokumentu, vytvořit ze zadaného dotazu frázi (bez nutnosti použití uvozovek) nebo pevně nastavit, aby název hledaného dokumentu začínal zadanými klíčovými slovy, a dále ještě upřesnit, ve kterých metadatových polích má systém vyhledávat. Přesná podoba úvodní obrazovky i primárního jednoduchého rozhraní závisí na nastavení každé knihovny, základem je však vždy jediné textové pole pro vyhledávání ve stylu vyhledávače Google.

Alternativou k jednoduchému vyhledávání je pokročilé vyhledávací rozhraní, které v případě systému Primo (v testovací implementaci UK) nenabízí rozsáhlé možnosti nastavení. Je v něm možné využít více dotazovacích polí a kombinovat dotazy v rámci více metadatových polí. Dále v něm lze upřesnit datum publikování dokumentu, a to i s přesností na dny a měsíce v roce zadáním přesného rozsahu do poskytnutých polí. Další možností upřesnění je podle typu materiálu: *knihy*, *články*, *seriály*, *obrázky* nebo *audiovizuální*. V poslední rolovací nabídce se skrývá nastavení jazyka, které skrývá pouze angličtinu, francouzštinu a němčinu (ale toto je pouze otázka nastavení). Nutno podotknout, že i pokročilé vyhledávací rozhraní lze nastavit na míru podle potřeb jednotlivých knihoven, a proto příklad z testovací verze UK není plně reprezentativní.

#### 2.3.4 Výsledky vyhledávání

Výsledky vyhledávání jsou v discovery systému Primo řazeny primárně podle relevance, dále je možné je řadit podle data (od nejnovějších), oblíbenosti a abecedně dle autora nebo názvu (viz obr. 6). Standardně se zobrazí prvních 10 záznamů.

The screenshot shows the Primo search interface. At the top, there's a search bar with the query 'web scale discovery'. Below it, there are filters for 'Všechny položky', 'obsahující slova mého dotazu', and 'kdekoliv v záznamu'. On the left, there's a sidebar with 'Rozšířit můj dotaz' (Expand my query) and 'Zobrazit pouze' (Show only) sections. The 'Zobrazit pouze' section has a dropdown menu with 'Recenzované časopisy (47 457)' and 'Dostupné (3)'. The 'Upřesnit výsledky' (Refine results) section has a list of categories: 'Typ zdroje' (Source type) with 'Články (52 820)', 'Novinové články (16 233)', 'Sborníky z konferencí (11 258)', 'Knihy (8 698)', 'Recenze (4 705)', and 'Další možnosti' (Other options). The 'Sbírka' (Collection) section has 'OneFile (GALE) (34 608)', 'IEEE Proceedings (11 092)', 'SciVerse ScienceDirect (Elsevier) (8 698)', 'Health Reference Center Academic (Gale) (6 473)', and 'MEDLINE (NLM) (5 173)'. The 'Téma' (Topic) section has 'Internet (2 765)', 'Web Services (1 526)', 'Semantic Web (597)', and 'Information Services (393)'. The main results area shows a list of search results. The first result is 'Coming into focus: Web scale discovery services face growing need for best practices.(DISCOVERY SERVICES)' by Kelley, Michael, Library Journal, Oct 15, 2012, Vol. 137(17), p.34(7). The second result is 'Web scale discovery services' by Jason Vaughan ebrary, Inc., Chicago, Ill. : American Library Association 2011. The third result is 'EXPLORE WEB SCALE DISCOVERY SERVICES IN NEW ISSUE OF LIBRARY TECHNOLOGY REPORTS' by States News Service, Feb 1, 2011. The fourth result is 'The discovery experience.(WEBCAST ALERT)(Defining Web-Scale Discovery: The Promise of a Unified Search Index for Libraries)(Brief article)' by Library Journal, Sept 15, 2009, Vol. 134(15), p.14(1). The fifth result is 'Web scale discovery: the user experience' by Julia Gross ; Lutie Sheridan. At the bottom right, there's a button 'Napište nám' (Write to us).

Obr. 6: Ukázka výsledků vyhledávání v systému Primo

Algoritmus pro výpočet relevance používaný v systému Primo se nazývá ScholarRank a je možné jej v několika ohledech přizpůsobit na míru potřebám knihovny. Zaprvé lze lokálním zdrojům knihovny nastavit vyšší důležitost a zobrazit je tím mezi prvními vyhledanými. Zadruhé je možné upravit relevanci pro všechny zdroje z celého indexu nastavením relativní váhy pro jednotlivá metadatová pole (název, autor, předmět atp.). Například při nastavení váhy 1.5 pro název a váhy 1.0 pro předmět jsou upřednostňovány ve výpisu výsledky obsahující dotaz přímo v názvu. V opačném případě by to byly primárně výsledky, jejichž předmětové zařazení odpovídá zadanému dotazu a přitom nemají nutně daná slova v názvu. Celý algoritmus je nicméně mnohem složitější a jeho přesná dokumentace není veřejně dostupná, také je neustále vylepšován a upravován ve snaze poskytovat uživateli co nejvíce relevantní výsledky na jeho dotazy. Primo také umožňuje personalizovat vyhledávání pro každého jednotlivého uživatele nastavením v osobním účtu (popsáno dále v textu).

Po levé straně od vyhledaných záznamů se nachází panel fasetové navigace, pomocí něž je možné upřesnit vyhledané výsledky. Nejprve je možné rozšířit hledání na dokumenty, ke kterým není dostupný plný text. Další kritéria již výsledný počet vyhledaných výsledků omezují.

Vyfiltrovat lze např. pouze dokumenty z recenzovaných časopisů. Pomocí fasetové kategorie *Typ zdroje* lze vypsát články, novinové články, sborníky z konferencí, knihy atd. Kategorie *Sbírka* umožní zobrazit dokumenty pouze z konkrétních zdrojových databází.

Kategorie *Téma* zastupuje v Primu předmětová hesla (která nejsou řízená – je tedy možné, že se jedno předmětové heslo zobrazí v několika různých variantách). Dalšími kritérii jsou jména autorů, datum vytvoření (bez možnosti zadat přesný rozsah s přesností na dny a měsíce, ale pouze upřesnit rozsah v řádu let), jazyk anebo název konkrétní seriálové publikace.

Kliknutím na odkaz *Další možnosti* umístěný pod každou kategorií se zobrazí nová obrazovka, na které lze zaškrtnout, které konkrétní možnosti do vyhledávání zahrnout, případně které naopak z vyhledávání vyřadit.

I nabídku fasetové navigace lze flexibilně přizpůsobit potřebám knihovny. Pro lokální zdroje je možné vytvořit zcela nové fasetové kategorie, které standardně systém nenabízí – například vedoucí práce či kategorie výpůjčky.

### **2.3.5 Zobrazení záznamu**

Výsledky vyhledávání v prostředí popsaném výše zobrazují zjednodušenou přehledovou formu záznamu. Pro každý typ obsahu existuje obrazová ikona, která se zobrazuje nalevo od textu záznamu a jednotlivé záznamy tak lze od sebe snáze odlišit.

V případě knih obsahuje záznam typicky následující druh údajů: název, autor a informace o publikaci (místo vydání, vydavatel a rok). Pokud si knihovna předplácí ještě nějakou službu pro obohacování obsahu<sup>15</sup>, zobrazí se místo standardní ikony přímo obrázek obálky. U knih fyzicky umístěných knihovně lze nastavit, jaká zpráva se zobrazí přímo v záznamu ohledně aktuálního stavu výpůjčky (např. „dostupné“, pokud je kniha volně dostupná k vypůjčení).

V případě článků z odborných časopisů obsahuje záznam typicky tyto údaje: název, autor, název časopisu, datum vydání, ročník a číslo, rozsah stran a poznámku, zdali jde o recenzovaný časopis. Může také obsahovat úryvek z plného textu nebo tzv. *Popis* – v něm může být zobrazen abstrakt, ale také anotace, referát nebo obsah z vybraných polí schémat MARC21 nebo Dublin Core (navenek není zřejmé, o který druh popisku se jedná).

Pod zkráceným záznamem kteréhokoliv typu zdroje se nachází několik odkazů pro získání rozšířených informací o dokumentu. Standardně jde o odkazy typu *Podrobnosti*, který uživatele přenesení na podrobnější formu záznamu obohacenou o popisky jednotlivých metadatových polí a několik dodatečných údajů (předmětová hesla, jazyk dokumentu, zdroj

---

<sup>15</sup> tzv. „enrichment services“; Primo podporuje například: Syndetic Solutions, Content Café nebo Amazon

záznamu atp.). V rámci tohoto detailu záznamu lze využít rozšířené funkcionality: uložit záznam do schránky, odeslat jej emailem, exportovat do citačních nástrojů EndNote a RefWorks anebo sociálních sítí pro vytváření záložek Connotea<sup>16</sup> a Del.icio.us. Dále *Recenze a značky*, kam mohou přihlášení uživatelé přidávat vlastní recenze a značky (tagy), a *Další služby*. U elektronicky dostupných dokumentů je typicky také uveden odkaz *Online zdroj* pro přístup přímo k plnému textu.

Při kliknutí na kterýkoliv z výše popsaných odkazů se nejprve rozbálí vnořené okno přímo pod záznamem v seznamu výsledků a až při kliknutí na ikonku v pravém horním rohu tohoto okna je uživatel přenesen na novou stránku se záznamem zobrazeným přes celou obrazovku. To samé se také může stát při kliknutí na nadpis záznamu, pokud není systém nastaven tak, aby uživatele odkázal přímo na zdroj plného textu, pokud je dostupný (takové je nastavení v případě testovací verze UK).

Přístup k plným textům je v discovery systému Primo řešen skrze dostupný odkazovací nástroj, tzv. OpenURL link resolver. Záleží na knihovně, který konkrétně má již zavedený, Primo podporuje řadu komerčních resolverů, přičemž sama společnost Ex Libris nabízí svůj nástroj SFX. Plné texty se tedy nezobrazují přímo v rozhraní Prima, ale link resolver uživatele navede do příslušné databáze, kde je plný text uložen.

### **2.3.6 Další funkce**

Jak již bylo popsáno dříve, vyhledávací vrstvu - systém Primo - lze použít i bez centrálního indexu Primo Central čistě jako OPAC nové generace. Proto nabízí také komplexní funkcionalitu knihovního katalogu. U dokumentů ze sbírek knihovny se tak v záznamu nachází ještě záložka „Exempláře“, pod kterou lze nalézt vypsaná umístění knihy v rámci knihoven UK, včetně signatury, přesného umístění a aktuální informace o stavu výpůjčky (zdali je dostupná, eventuálně dokdy je vypůjčená). Pomocí záložky „Exempláře“ lze přímo z Prima zadat požadavek na výpůjčku (rezervaci) žádaného dokumentu (tato funkce je dostupná až po přihlášení do čtenářského konta).

Díky katalogovým funkcím systému Primo lze přímo z něj také pohodlně spravovat své čtenářské konto. Přístup do něj je skrze odkaz „Můj účet“ v pravé horní části stránky. Ve správě účtu čtenář nalezne svůj aktuální seznam výpůjček se všemi údaji a možností výpůjčky

---

<sup>16</sup> Služba Connotea k 12. březnu 2013 ukončila svůj provoz.

prodloužit. Může zde také spravovat své rezervace, pokuty a poplatky a editovat své osobní údaje.

V rozhraní *Schránka*, dostupném přes odkaz v pravé horní části kterékoli stránky, lze pracovat se záznamy, které si uživatel během vyhledávání uložil (např. označením hvězdičkou). Je možné si je řadit do složek a složky i samotné záznamy opatřit vlastní poznámkou a opět využít funkcí jako: odeslat emailem, exportovat do již zmíněných služeb EndNote, RefWorks, Connotea a Del.icio.us anebo záznam odeslat k tisku. V rozhraní *Schránky* se označený záznam vždy v kompletním znění zobrazuje vespod stránky.

V nejnovější verzi systému Primo přibyla ještě možnost v účtu nastavit *Personalizaci vyhledávání*. Uživatel zvolí svou úroveň (např. student, výzkumník nebo knihovník) a obor (je možné zatrhnout libovolné množství oborů). Na základě toho jsou mu při vyhledávání nabízeny výsledky nejen na základě běžného algoritmu pro výpočet relevance, ale je přiřazena vyšší důležitost takovým zdrojům, o kterých si systém myslí, že jsou pro daného uživatele relevantnější, protože např. odpovídají zvolenému oboru.

### **2.3.7 Shrnutí**

Web scale discovery řešení společnosti Ex Libris těží nejen z rozsáhlého centrálního indexu Primo Central, ale také discovery vrstvy Primo, která zprostředkovává plnohodnotné funkce knihovního katalogu nové generace a tím významně odlišuje kompletní systém od konkurence představované v této práci. Velkou výhodou je plné využití stávající registrace uživatelů.

Práce se systémem je dostatečně intuitivní pro běžné návštěvníky knihoven, ale systém nabízí také řadu funkcí pro zpřesnění dotazu, které uspokojí i uživatele se specifitějšími potřebami a knihovníky.

## **2.4 Summon**

### **2.4.1 Charakteristika systému**

Web scale discovery systém Summon<sup>17</sup> společnosti Serials Solutions vstoupil na trh v červenci 2009. Jeho vývoj započal v roce 2008 a systém byl od počátku koncipován a vyvíjen jako zcela nová platforma. Zákazníkům je nabízen vzdáleně způsobem software jako služba a součástí produktu je vyhledávací systém i rozsáhlý centrální index.

---

<sup>17</sup> Oficiální webová stránka - <http://www.serialssolutions.com/en/services/summon/>

V současné době si předplácí web scale discovery systém Summon více než 550 institucí po celém světě. V České republice jsou v současné době provozovány tři placené instalace. Hlavními předplatiteli jsou především knihovny akademické a odborné, dále veřejné, knihovny vládních institucí a soukromé společnosti.

Zprovoznění systému od finálního podepsání objednávky trvá okolo 8-12 týdnů, nejvíce času je potřeba na implementaci OPACu a repozitářů knihovny do centrálního indexu. Cena systému se odvíjí od velikosti a typu instituce (u vysokoškolských knihoven například podle počtu studentů a odborných pracovníků). Aktualizace systému přicházejí v podstatě neustále, zhruba každý týden. Větší updaty vycházejí několikrát do roka.

V České republice distribuuje Summon společnost Alberta icome Praha s.r.o., která také poskytuje knihovnám technickou podporu při implementaci i následně v době provozu systému<sup>18</sup>.

#### **2.4.2 Obsah**

Discovery systém Summon využívá jeden velký vzdáleně hostovaný centrální index, který sdružuje jak obsah získaný licenčními dohodami s poskytovateli obsahu tak lokální zdroje knihoven.

Summon dokáže sklídit lokální obsah ze všech standardních katalogů a repozitářů, podporuje běžná datová schémata MARC, Dublin Core, XML nebo EAD a metody doručování obsahu OAI-PMT a FTP. Lokální obsah je nahrán do centrálního indexu a všichni uživatelé Summonu prohledávají jeho kompletní šíři. Každá knihovna se však může rozhodnout, že její obsah veřejně dostupný nebude (Vaughan, 2011c).

Celková velikost centrálního indexu Summon je v současnosti přes 1,2 miliardy dokumentů, z nichž část není veřejně přístupná (viz výše). Summon nabízí záznamy z odborných zdrojů získané dohodami s mnoha významnými vydavateli, agregátory i open-access repozitáři a kromě toho také velké množství novinových článků<sup>19</sup>.

---

<sup>18</sup> Necitované informace v kapitolách *Charakteristika systému* a *Obsah* poskytl Ing. Vladimír Karen, výkonný ředitel firmy Alberta icome Praha, při emailové komunikaci.

<sup>19</sup> Na webových stránkách - <http://www.serialssolutions.com/en/services/summon/content> – jsou k dispozici rozsáhlé seznamy participujících vydavatelů a konkrétní míry pokrytí klíčových databází.

### 2.4.3 Možnosti vyhledávání

Použití web scale discovery systému Summon bude představeno na implementaci Národní technické knihovny<sup>20</sup> (viz obr. 7).



Obr. 7: Úvodní stránka systému Summon

Pro přístup do systému Summon a k prohledávání jeho kompletního centrálního indexu není vyžadováno přihlášení. To je potřeba až v momentu pokusu o přístup k plnému textu dokumentu (Vaughan, 2011c).

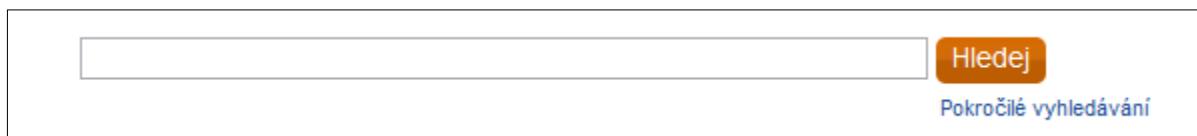
Uživatelské rozhraní web scale discovery systému Summon obsahuje kromě základního prostoru pro vyhledávání ovládací panel ve svrchní části obrazovky a vedlejší pomocný panel vespod. Vrchní panel odkazuje na nápovědu systému, krátký propagační text o funkcích systému, kontaktní formulář a nastavení jazyka. Spodní panel během samotného vyhledávání odkazuje především na uživatelem uložené záznamy.

K vyhledávání slouží v případě discovery systému Summon primárně jednoduché základní vyhledávací rozhraní skládající se z jediného textového pole a aktivního tlačítka *Hledej* a odkazu na pokročilou verzi vyhledávacího rozhraní (viz obr. 8). Slouží k zadání dotazu na základě klíčových slov, s možností omezení přesné fráze pomocí uvozovek a použití logických booleovských operátorů: OR, NOT a AND (zapsaných velkými písmeny).

<sup>20</sup> <http://techlib.summon.serialssolutions.com/>



Lze též využít zástupných znaků „?“ a „\*“ pro nahrazení jednoho, respektive více znaků ve vyhledávaném termínu.

The image shows a search interface for the Summon system. It features a long, empty search input field on the left. To the right of the input field is an orange button with the white text "Hledej". Below the button, the text "Pokročilé vyhledávání" is displayed in a smaller, blue font.

Obr. 8: Základní vyhledávací rozhraní systému Summon

V případě potřeby je možné využít *Pokročilé vyhledávání*, které umožňuje bližší specifikaci vyhledávacího dotazu. Kromě klíčových slov lze vyhledávat i v polích: autor, název, ISBN, ISSN, zdrojový dokument, ročník, číslo, datum vydání a typ obsahu. Dále je možné již předem omezit vyhledávání pouze na: *záznamy s plným textem dostupným online; odborné vědecké materiály, včetně recenzovaných; recenzované publikace; záznamy v katalogu knihovny (převážně tištěné a fyzické materiály)* anebo na kombinaci těchto kritérií. Stejným způsobem je také možné předem z vyhledávání vyloučit: *novinové články; knižní recenze; dizertace / závěrečné práce*. Poslední zaškrťovací okénko umožní předem rozšířit vyhledávání o záznamy mimo fond lokálních a předplacených zdrojů místní knihovny (tedy zahrnout kompletní veřejně dostupný obsah centrálního indexu).

#### 2.4.4 Výsledky vyhledávání

Výsledky vyhledávání jsou v discovery systému Summon řazeny primárně podle relevance, alternativně podle data (od nejnovějších nebo od nejstarších). Standardně se zobrazí prvních 10 záznamů (viz obr. 9).

NTK  
50° 14' 02" N, 14° 22' 26.385" E  
Národní technická knihovna  
National Technical Library

Web NTK | Nápověda | O Summonu | Kontaktujte nás | Čeština

web scale discovery Hledej Pokročilé vyhledávání

☐ Zachovat nastavení vyhledávání ☒ Nové vyhledávání

Výsledky vyhledávání: Vašemu dotazu **web scale discovery** odpovídá **67,811** výsledků

**Upřesnění dotazu**

- ☐ Záznamy s plným textem dostupným online
- ☐ Pouze články z odborných vědeckých časopisů, včetně recenzovaných
- ☐ Pouze články z recenzovaných publikací
- ☐ Vyloučit novinové články
- ☐ Záznamy v katalogu knihovny
- ☐ Rozšířit vyhledávání i mimo Vaši knihovnu

▼ **Typ obsahu**

- ☒ Vše
- ☐ Článek v odborném časopise (65,080)
- ☐ Novinový článek (2,154)
- ☐ Knižní recenze (573)
- ☐ Dizertace / Závěrečné práce (493)
- ☐ Konferenční sborník (426)
- ☐ Kniha / eBook (297)
- [další...](#)

▼ **Předmětová hesla**

- ☒ Vše
- ☐ studies (10,773)
- ☐ research (5,784)
- ☐ psychology (4,844)
- ☐ scale (3,835)
- ☐ environmental sciences

**Doporučení:** Našli jsme jednu nebo více specializovaných kolekcí, které by Vás mohly zajímat.

- Web of Knowledge** - Covers over 10,000 of the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings

**Web-Scale Discovery**

autor Jason Vaughan  
American Libraries, ISSN 0002-9769, 01/2011, Ročník 42, Číslo 1/2, p. 32  
...Connecting users with the information they seek is one of the central pillars of our profession. **Web-scale discovery** services for libraries are those services... Libraries, Research  
[Článek v odborném časopise: Plný text online](#)

**Investigations into library web-scale discovery services**

autor Vaughan, Jason  
Information Technology and Libraries, ISSN 0730-9295, 03/2012, Ročník 31, Číslo 1, p. 32  
...ABSTRACT **Web-scale discovery** services for libraries provide deep **discovery** to a library's local and licensed content and represent an evolution--perhaps... Web 2.0, Algorithms, Research, Libraries and readers, Collection development (Libraries)  
[Článek v odborném časopise: Plný text online](#)

**Web scale discovery: the user experience**

autor Julia Gross a Lutie Sheridan  
New Library World, ISSN 0307-4803, 2011, Ročník 112, Číslo 5/6, pp. 236 - 247  
Purpose – This study aims to look at how a small group of university students used the new library **web scale search discovery** tool, "Summon... Web sites, Information retrieval, Telecommunications systems & Internet communications, Experimental/theoretical, Asia & the Pacific  
[Článek v odborném časopise: Plný text online](#)

**Web Scale Discovery Services: A Library Technology Report**

autor Vaughan, Jason

2013 Národní technická knihovna | S podporou Summon™ Uložené záznamy (0)

Obr. 9: Ukázka výsledků vyhledávání v systému Summon

Summon využívá vlastní algoritmus pro výpočet relevance, který prohledává metadata i plný text dokumentu na základě zadaných klíčových slov. Různá pole metadat mají pro výslednou relevanci rozdílnou důležitost. Například v případě článku v časopise se relevance určuje mj. podle toho, zdali článek vyšel v recenzovaném časopise a kolikrát byl citován (Vaughan, 2011c, s. 23). Nastavení algoritmu relevance může knihovna ovlivnit dohodou se společností Serials Solutions – lze např. více preferovat lokální zdroje z katalogu knihovny anebo články z odborných periodik vůči novinovým. Kompletní dokumentace algoritmu není veřejně dostupná a samotný algoritmus je také neustále vylepšován.

Po levé straně od vyhledaných záznamů se nachází panel fasetové navigace, pomocí níž je možné dodatečně upřesnit vyhledané výsledky. První kategorie *Upřesnění dotazu* sestává ze stejných upřesňujících kritérií jaké je již možné zaškrtnout předem v pokročilém vyhledávacím rozhraní.

Druhá kategorie *Typ obsahu* umožňuje omezit výsledky pouze na vybrané konkrétní typy dokumentů typu: článek v odborném časopise, novinový článek, dizertace / závěrečná

práce, knižní recenze, knihy / eknihy atp. Po kliknutí na odkaz *další...* je možné vybírat z kompletního seznamu typů obsahu a vybrat konkrétně, který typ zařadit a který naopak z výsledků vyhledávání vyřadit.

Třetí kategorií jsou *Předmětová hesla*. Ve fasetě se zobrazuje zhruba 100 nejfrekventovanějších hesel z vyhledaných záznamů a protože nejsou řízená, mohou se mezi nimi vyskytnout duplicity stejného hesla v různých podobách.

Dalším fasetovým kritériem je omezení podle data, a to buď za pomoci posuvníku a nastavení data publikace dokumentu s přesností na roky anebo vepsání přesného rozmezí s přesností na den.

Poslední možností filtrování výsledků je podle jazyka.

Nad seznamem vyhledaných záznamů se zobrazuje panel *Doporučení*, kde jsou uživateli navrhovány celé konkrétní databáze relevantní pro jeho zadaný dotaz a do kterých se může přenést skrze přímý odkaz.

Pod seznamem vyhledaných záznamů se zobrazuje seznam návrhů na zpřesnění zadaného dotazu. Například při vyhledání termínu „web scale“ je nabídnuto „web scale discovery“ a „web scale discovery services“. Pokud by byl zadaný termín např. „Psychology“, bylo by jich mnohem více.

#### **2.4.5 Zobrazení záznamu**

Výsledky vyhledávání v prostředí popsaném výše standardně zobrazují zjednodušenou formu záznamu. Každý typ obsahu má svou vlastní obrazovou ikonu, která se zobrazuje nalevo od textu záznamu. Pokud ji překrývá výrazně žlutá paprskovitě oválná ikona s nápisem „online“, znamená to, že je dokument dostupný v plném textu.

K plnému textu se uživatel standardně dostane kliknutím na název záznamu. Přístup k plným textům je v discovery systému Summon řešen skrze dostupný linkovací nástroj. Národní technická knihovna využívá systém SFX společnosti Ex Libris.

V případě knih obsahuje záznam typicky následující druh údajů: název, autor, datum vydání, ISBN, počet stran a předmětová hesla. U knih dostupných fyzicky v knihovně se zobrazuje ještě aktuální údaj o jejich dostupnosti a lokaci. U elektronických knih je na stejném místě odkaz na plný text. U některých knih se zobrazuje místo standardní knižní ikony pro typ obsahu kniha přímo obálka dané knihy – tato funkcionalita záleží na smluvních

licenčních dohodách, které má knihovna nastavené ohledně poskytovatelů doplňkového obohacujícího obsahu, a dostupnosti konkrétních obrazových materiálů (Vaughan, 2011c).

U dokumentů typu článek v odborném časopise, novinový článek anebo zpravodaj obsahuje záznam typicky následující údaje: název, autor, název publikace, ISSN, datum vydání, ročník a číslo. Mohou také obsahovat několik řádků z plného textu, předmětová hesla, údaj o stránkách od-do, anebo údaj *Od strany*, značící na kolikáté straně zdrojového dokumentu vyhledaný text začíná (nikoliv počet stran).

U kteréhokoli typu dokumentu se při najetí kurzorem myši na název (anebo kliknutím na ikonu lupy v pravém horním rohu záznamu) zobrazí detailní náhled záznamu obsahující údaje již zobrazené ve zjednodušeném výpisu, ale i s popisky jednotlivých polí a eventuelně dalšími údaji typu: abstrakt, jazyk dokumentu, žánr (např. interview, news) a údaj o majiteli copyrightu.

Kliknutím na ikonku složky se znakem plus je možné záznam přidat do uložených záznamů. Z těch pak lze vytvářet citace podle mezinárodních norem, exportovat do citačních manažerů, poslat na email nebo vytisknout. Uložené záznamy se ve složce uchovávají pouze krátkodobě – jen do té doby, než uživatel ukončí práci s internetovým prohlížečem.

#### **2.4.6 Další funkce**

Summon nenabízí žádnou možnost vytvoření si uživatelského účtu, která by obsahovala například možnosti úpravy uživatelského rozhraní nebo dlouhodobou práci s uloženými záznamy. Případné přihlášení do účtu instituce (pomocí odkazu v šedé horní liště) slouží pouze k přímému přístupu k plným textům v rámci dané relace prohlížeče.

Summon také neposkytuje žádné další funkce knihovního katalogu kromě zobrazení lokace dokumentu. Akce typu prodloužení výpůjčky nebo správu svého knihovního konta musí uživatel řešit zvlášť v katalogu.

#### **2.4.7 Shrnutí**

Web scale discovery systém Summon klade velký důraz na skutečnou jednoduchost vyhledávání a uživatelskou přívětivost při práci se systémem. Uživatelské rozhraní je jednoduché a graficky čisté v duchu vyhledávače Google a nechybí mu klasické prvky discovery systémů, jako je fasetová navigace. Záznamy dokumentů obsahují základní informace a nejsou přespříliš komplikované.

Summon těží z ohromného centrálního indexu poskytujícího velké množství výsledků na zadané dotazy, ale je to též dáno např. zařazením velkého množství novinových článků, jež nespádají pod odborné zdroje (v případě potřeby je lze jednoduše vyfiltrovat).

Za možnou nevýhodu lze považovat takřka nulové využití stávající, nebo jakékoli jiné, registrace uživatelů a z toho vyplývající neexistující dodatečné možnosti práce se systémem (např. ukládání výsledků a dlouhodobější práce se záznamy). Summon také nenabízí žádné funkce knihovního katalogu kromě zobrazení aktuálního stavu a lokace dokumentu.

### 3 Srovnání vybraných systémů z hlediska funkčních možností

V této závěrečné kapitole bude představeno srovnání výše představených systémů z hlediska funkčních možností. Cílem není porovnat každou jednotlivou nabízenou funkci, nýbrž všeobecné funkční vlastnosti, kterými se od sebe jednotlivá discovery řešení mohou odlišovat a která hrají zásadní roli v rozhodování knihovny při výběrání systému ke koupi.

Srovnání je rozděleno do několika kategorií podle jednotlivých funkčních oblastí systémů, přičemž u každé jednotlivé vlastnosti je uvedeno **Ano** nebo **Ne** podle toho, zdali ji systém nabízí, nebo ne. Podle potřeby je připsáno dodatečné slovní vysvětlení. Pod každou tabulkou srovnání pak následuje ještě slovní doplnění k porovnávaným vlastnostem.

#### 3.1 Obecné funkce systému

	<b>EDS</b>	<b>Primo</b>	<b>Summon</b>
Plný přístup bez přihlášení	<b>Ne</b> Pouze tzv. „Guest“	<b>Ano</b>	<b>Ano</b>
Využití instituční registrace uživatele	<b>Ne</b> Pro vybrané funkce nutnost založení bezplatného účtu My EBSCOhost	<b>Ano</b>	<b>Ano</b> Ale pouze zjednoduší přístup k plným textům
Funkce knihovního katalogu	<b>Ne</b>	<b>Ano</b>	<b>Ne</b>
Integrace lokálních zdrojů knihovny do jednotného vyhledávání	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Mobilní verze stránek	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Aplikace pro mobilní telefony	<b>Ano</b> iOS, Android	<b>Ne</b>	<b>Ne</b>

Tabulka 1: Porovnání obecných funkcí systémů

Největším rozdíly mezi analyzovanými systémy jsou již v základních vlastnostech – přístup, využití existující institucionální registrace uživatele a funkce knihovního katalogu. Zatímco Primo a Summon nevyžadují přihlášení až do doby, kdy se uživatel pokusí přistoupit k plnému textu, u EDS je vyhledávání bez přihlášení značně omezené (především z hlediska zobrazovaného obsahu). Jediné Primo dokáže plně využít stávající registraci a navíc nabízí plnohodnotné zastoupení knihovního katalogu. Přestože tato funkcionalita nepatří do definice web scale discovery systémů jako takových, pro knihovnu může hrát při rozhodování významnou roli, obzvláště pokud současně využívaný katalog je již zastaralý a nesplňuje současné požadavky.

### 3.2 Funkce vyhledávání

	<b>EDS</b>	<b>Primo</b>	<b>Summon</b>
Jednoduché vyhledávání	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Pokročilé vyhledávání	<b>Ano</b> Dále také vizuální	<b>Ano</b>	<b>Ano</b>
Fasetová navigace	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Našeptávač	<b>Ano</b>	<b>Ne</b>	<b>Ano</b>
Oprava překlepů	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Fráze, zástupné znaky a booleovské operátory	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Historie vyhledávaných dotazů	<b>Ano</b> Krátkodobá (pouze po dobu jedné relace), po přihlášení možnost dlouhodobého uložení dotazu	<b>Ano</b> Krátkodobá (pouze po dobu jedné relace), po přihlášení možnost dlouhodobého uložení dotazu	<b>Ne</b>

Tabulka 2: Porovnání funkcí vyhledávání

Snahou web scale discovery systémů je co nejvíce vyhledávací proces zjednodušit tak, aby se k relevantním výsledkům dostal i uživatel bez jakýchkoli odborných znalostí zvyklý na moderní internetové prostředí, jemuž na poli vyhledávačů vévodí Google. Základní vyhledávací rozhraní skládající se z jednoho políčka nabízejí všechny analyzované systémy bez rozdílu, žádný se také „neodvážil“ vynechat rozhraní pokročilé nebo podporu booleovských operátorů. Důraz je ale kladen na fasetovou navigaci a zpřesnění dotazu až po provedení vyhledávání.

### 3.3 Záznamy

	EDS	Primo	Summon
Základní bibliografické údaje	Ano	Ano	Ano
Abstrakt	Ano	Ano V rámci pole <i>Popis</i>	Ano
Předmětová hesla	Ano	Ano	Ano
Údaj o zdroji	Ano	Ano	Ne
Alertové služby	Ano RSS, email	Ano RSS, email	Ano RSS
Údaj o aktuální dostupnosti dokumentu (u lokálních zdrojů)	Ano	Ano	Ano
Obohacení záznamů o data z externích zdrojů	Ano	Ano	Ano
Ukládání záznamů do schránky	Ano Po přihlášení i dlouhodobě	Ano Po přihlášení i dlouhodobě	Ano Pouze krátkodobě po dobu práce se systémem
Práce se záznamy (email, tisk, export, citace)	Ano	Ano Kromě přímé tvorby citací	Ano

Tabulka 3: Porovnání zobrazovaných záznamů a jejich funkcí

Mít co nejkvalitnější záznamy s bohatými metadaty je cílem každého producenta web scale discovery systémů a jejich centrálních indexů. Systémy sklízí metadata, eventuelně i plné texty, z licenčně nasmlouvaných databází, avšak každý discovery systém používá vlastní schéma záznamu a ne vždy zobrazuje všechna metadata, která jsou jinak ve zdrojové databázi dostupná. Proto se od sebe jednotlivé systémy a jejich záznamy v několika ohledech odlišují.

Předmětová hesla nejsou vzhledem k množství indexovaného obsahu v žádném ze systému 100% řízená. V systému EDS záleží jejich kvalita výrazně na tom, jestli záznam pochází z vlastních databází EBSCO, do záznamů v systému Primo lze zakomponovat data z tezaurů, pokud je má knihovna k dispozici, a systém Summon třídí záznamy do 59 vědních oborů podle klasifikace HILCC, ale v současné verzi uživatel toto třídění nemůže prakticky využít.

A systémy se liší nejen mírou zobrazovaných metadat, ale také nabízenými funkcemi pro práci se záznamy – Summon například neumožňuje dlouhodobé uložení záznamů<sup>21</sup>.

<sup>21</sup> V chystané nové verzi systému Summon 2.0 to již možné bude



Všechny systémy ale podporují minimálně některý způsob vytváření alertů, dokáží obohatit záznamy o data z externích zdrojů (např. obrázky knižních obálek) a podporují další práci se záznamy.

### 3.4 Možnosti nastavení systému knihovnou

	<b>EDS</b>	<b>Primo</b>	<b>Summon</b>
Ovlivnění algoritmu relevance vyhledávání	<b>Ano</b> Pouze u lokálních zdrojů	<b>Ano</b>	<b>Ano</b> Dohodou s producentem systému
Možnost upravit vzhled systému	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>
Možnost upravit vybrané funkce systému	<b>Ano</b>	<b>Ano</b>	<b>Ano</b>

**Tabulka 4: Porovnání možností nastavení systémů knihovnou**

Všechny popisované systémy umožňují do jisté míry knihovnám ovlivnit algoritmus relevance vyhledávání. Někdy lze změny provádět přímo v administračním systému, u systému Summon pak např. pouze zprostředkovaně. Míra nastavení se liší systém od systému. Všechny nabízejí úpravu důležitosti lokálních zdrojů oproti výsledkům z celého centrálního indexu, Primo a Summon i určitá nastavení týkající se všech výsledků. Systém Primo nabízí možnost ovlivnit kompletní algoritmus pro všechny výsledky nastavením vah pro jednotlivá metadatová pole (název, autor, předmět atp.), systém Summon lze nastavit např. tak, aby preferoval odborné články oproti novinovým.

Nastavení vzhledu a jednotlivých funkcí je pro knihovnu zcela bezpochyby významným faktorem při rozhodování o koupi, v rámci této práce (psané z pohledu studenta-uživatele) však nebylo možné – bez přístupu do administračních rozhraní jednotlivých systémů a v rámci stanoveného rozsahu práce – jít více do hloubky. Všechny z analyzovaných systémů lze z hlediska vzhledu alespoň základně přizpůsobit (logo knihovny, barevná šablona, vyhledávací okénko na úvodní stránce knihovny apod.) a všechny nabízí také minimálně několik možností ovlivnit funkční chování (např. jestli je po kliknutí na nadpis záznamu uživatel odkázán na detailní záznam anebo přímo na plný text).

### 3.5 Možnosti nastavení systému uživatelem

	<b>EDS</b>	<b>Primo</b>	<b>Summon</b>
Úpravy vzhledu systému	<b>Ano</b>	<b>Ne</b>	<b>Ne</b>
Ovlivnění algoritmu relevance vyhledávání	<b>Ne</b>	<b>Ano</b> Nastavením své úrovně a preferovaných oborů	<b>Ne</b>

Tabulka 5: Porovnání možností nastavení systémů uživatelem

Poskytnout uživatelům možnost upravit si systém podle svých potřeb se nezdá být, na rozdíl od (grafických i funkčních) úprav podle potřeb instituce, prioritou producentů. Analyzované systémy příliš možností osobního nastavení neposkytují. Výjimkou je možnost měnit uspořádání uživatelského rozhraní v systému EDS a možnost lehce ovlivnit algoritmus relevance vyhledávání v případě systému Primo podle odborného zaměření uživatele.

## 4 Závěr

Cílem bakalářské práce bylo představit knihovní vyhledávací systémy nové generace, které se snaží o to samé, co se již klasickým vyhledávačům, jako je Google, podařilo – přinášet uživatelům odpovědi na jejich dotazy rychle, přesně a jednoduše. Tyto nové vyhledávací systémy, pro které se zrodilo označení web scale discovery, přináší okamžité výsledky díky sdružení různých zdrojů do jediného centrální indexu, větší uživatelskou přívětivost díky důrazu na jednoduché vyhledávací rozhraní a používání prostých klíčových slov, ale to všechno s důrazem na odborné zdroje, které si knihovny předplácejí a chtějí dostat do co největšího povědomí svých uživatelům.

Nejprve byly vysvětleny některé druhy motivace, proč by se knihovny měly o novinky ve vyhledávání zajímat: kromě usnadnění a zpříjemnění práce uživatelům a snaha vést je k používání odborných zdrojů místo ne vždy ověřitelných informací z volného internetu je jedním z zásadních důvodů také zajištění efektivnějšího využívání předplácených elektronických zdrojů, na které knihovny vynakládají nemalé prostředky. Je představeno, jaké řešení přináší web scale discovery systémy, jež jsou různými autory prozatím definovány jako flexibilní systémy, jejichž fungování je založeno na obsahu uloženém ve velkém virtuálním skladu (centrální indexu), který je rychle prohledáván a výsledky jsou tříděny a poskytovány uživateli (pomocí tzv. discovery layer). Tvorba centrálních indexů je z velké části závislá na metadatech a plných textech poskytnutých poskytovateli obsahu v rámci dohod uzavíraných se vzájemně konkurujícími si producenty systémů, a právě proto se obsah a pokrytí jednotlivých vědních oborů a odborných titulů mezi jednotlivými systémy a jejich centrálními indexy mnohdy liší.

Popisovány a analyzovány jsou tři vybrané komerční web scale discovery systémy, které dohromady tvoří kompletní nabídku na českém knihovním trhu. Každý z nich je obecně charakterizován a následně analyzován s důrazem na funkční kapacity v jednotlivých kategoriích – vyhledávání, zobrazení výsledků, zobrazení záznamu a případné další funkce. Pro praktické testování byla zvolena u každého systému jedna fungující implementace systému.

Závěr práce je zaměřen na přehledné srovnání jednotlivých funkcí s cílem představit společné a odlišné prvky analyzovaných systémů. Navzdory tomu, že všechny splňují stejnou definici, najdou se mezi nimi rozdíly. Je nutné dodat, že není možné nyní prohlásit, který z představených systémů je ten nejlepší. Všechny mají své silné i slabší stránky a stejně jako

u výběru jakéhokoliv jiného knihovního systému záleží na celé řadě faktorů, z nichž podstatná je samozřejmě i cena, jakou producent nabídne.

Přestože bylo nad možností bakalářské práce provádět navíc ještě profesionálním způsobem výzkum použitelnosti testovaných systémů a zde se tak nabízí příležitost pro další autory s profesním nebo osobním zájmem o web scale discovery, je nutné na závěr dodat, že myšlenka představovaných vyhledávačů nové generace je určitě správná. Vyhledávání je zpravidla jednoduché a uživatelsky přívětivé, množství zobrazených výsledků ohromné a výsledky na první pohled relevantní. Právě na zdokonalování algoritmů relevance a co nejlepší zpracovávání uživateli zadávaných dotazů (třeba i v naprosto přirozeném jazyce) bude nyní u producentů web scale discovery služeb jistě kladen velký důraz, stejně tak jako je na ně kladen ve světě klasických internetových vyhledávačů. Uživatelé knihoven, kterým jde v zásadě o jediné – získat relevantní odpověď na svůj zadaný dotaz - z tohoto přístupu mohou jediné získat.

## Seznam použité literatury

- BATES, Marcia J. Improving User Access to Library Catalog and Portal Information. [online]. June 1, 2003 [cit. 2013-04-16]. Dostupný z: <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>
- BREEDING, Marshall. Library Web-Scale. *Computers in Libraries*. 2012a, **32**(1), 19-21. ISSN 1041-7915. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- BREEDING, Marshall. Looking Forward to the Next Generation of Discovery Services. *Computers in Libraries*. 2012b, **32**(2), 28-31. ISSN 1041-7915. Dostupný také komerčně ze systému EBSCOhost.com
- COUFALOVÁ, Jindřiška, Karolína KOŠŤÁLOVÁ a Hana NEMEŠKALOVÁ. *Katalogy nové generace: analýza vybraných systémů z pohledu uživatele*. Praha: Národní knihovna České republiky, 2009, 66 s. ISBN 978-80-7050-579-3.
- COUFALOVÁ, Jindřiška, Karolína KOŠŤÁLOVÁ a Petra ŠŤASTNÁ. Zkušenosti služeb Národní knihovny s centrálními indexy a web-scale discovery systémy. *Knihovna plus* [online]. 2012, 1 [cit. 2013-04-13]. ISSN 1802-8772. Dostupný z: <http://knihovna.nkp.cz/knihovnaplus121/coufal.htm>
- DEMPSEY, Lorcan. Web scale. In: *Lorcan Dempsey's Weblog: On libraries, services and networks* [online]. January 5, 2007 [cit. 2013-04-16]. Dostupný z: <http://orweblog.oclc.org/archives/001238.html>
- EBSCO Publishing. *EBSCO Discovery Service / EBSCOhost* [online]. © 2012 [cit. 8. ledna 2013]. Dostupný z: <http://www.ebscohost.com/discovery>
- Ex Libris Ltd. *Primo: Empowering Libraries to Address User Needs* [online]. © 2012 [cit. 30. dubna 2013]. Dostupný z: <http://www.exlibrisgroup.com/category/PrimoOverview>
- HOEPPNER, Athena. The Ins and Outs of Evaluating Web-scale Discovery Services. *Computers in Libraries*. 2012, **32**(3), 6-13. ISSN 1041-7915. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- KTD: *Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna České republiky, © 2009 [cit. 12. února 2013]. Dostupná z: [sigma.nkp.cz/cze/ktl](http://sigma.nkp.cz/cze/ktl)

- KELLEY, Michael. Stakeholders Strive to Define Standards for Web-Scale Discovery Systems. *The Digital Shift* [online]. October 11, 2012 [cit. 2013-04-16]. Dostupný z: <http://www.thedigitalshift.com/2012/10/discovery/coming-into-focus-web-scale-discovery-services-face-growing-need-for-best-practices/>
- KREJČÍŘ, Vlastimil. Vyhledávání v elektronických informačních zdrojích. *Zpravodaj ÚVT MU*. 2011, **21**(5). ISSN 1212-0901. Dostupný také z: <http://www.ics.muni.cz/bulletin/articles/676.html>
- KUČEROVÁ, Helena. Metavyhledávání a jeho typy – příspěvek k terminologické diskusi. *Knihovna plus* [online]. 2011, 2 [cit. 2013-04-16]. ISSN 1801-5948. Dostupný z: <http://knihovna.nkp.cz/knihovnaplus112/kucer.htm>
- National Information Standards Organization. NISO RP-19-201x Open Discovery Initiative (ODI). *NISO: How the information world connects* [online]. © 2013 [cit. 2013-04-16]. Dostupný z: [http://www.niso.org/apps/group\\_public/project/details.php?project\\_id=113](http://www.niso.org/apps/group_public/project/details.php?project_id=113)
- Serials Solutions. *The Summon® Service* [online]. © 2013a [cit. 8. ledna 2013]. Dostupný z: <http://www.serialssolutions.com/en/services/summon/>
- Serials Solutions. Summon help [online]. © 2013b [cit. 8. ledna 2013]. Dostupný z: <http://www.serialssolutions.com/en/services/summon/>
- VAUGHAN, Jason. Chapter 1: Web Scale Discovery What and Why?. *Library Technology Reports*. 2011a, **47**(1), 5-11. ISSN 0024-2586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- VAUGHAN, Jason. Chapter 2: OCLC WorldCat Local. *Library Technology Reports*. 2011b, **47**(1), 12-21. ISSN 0024-2586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- VAUGHAN, Jason. Chapter 3: Serials Solutions Summon. *Library Technology Reports*. 2011c, **47**(1), 22-29. ISSN 0024-2586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- VAUGHAN, Jason. Chapter 4: Ebsco Discovery Services. *Library Technology Reports*. 2011d, **47**(1), 30-38. ISSN 0024-2586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>

- VAUGHAN, Jason. Chapter 5: Ex Libris Primo Central. *Library Technology Reports*. 2011e, **47**(1), 39-47. ISSN 00242586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- VAUGHAN, Jason. Chapter 6: Differentiators and A Final Note. *Library Technology Reports*. 2011f, **47**(1), 48-53. ISSN 0024-2586. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>
- VAUGHAN, Jason. Investigations into Library Web-Scale Discovery Services. *Information Technology & Libraries*. 2012, **31**(1), 32-82. ISSN 0730-9295. Dostupný také komerčně ze systému EBSCOhost: <http://www.ebscohost.com>

## Seznam obrázků a tabulek

Obr. 1: Úvodní stránka systému EDS .....	18
Obr. 2: Základní vyhledávací rozhraní systému EDS.....	19
Obr. 3: Ukázka výsledků vyhledávání v systému EDS .....	20
Obr. 4: Úvodní stránka systému Primo.....	25
Obr. 5: Základní vyhledávací rozhraní systému Primo.....	26
Obr. 6: Ukázka výsledků vyhledávání v systému Primo .....	27
Obr. 7: Úvodní stránka systému Summon .....	32
Obr. 8: Základní vyhledávací rozhraní systému Summon.....	33
Obr. 9: Ukázka výsledků vyhledávání v systému Summon.....	34
Tabulka 1: Porovnání obecných funkcí systémů .....	38
Tabulka 2: Porovnání funkcí vyhledávání .....	39
Tabulka 3: Porovnání zobrazovaných záznamů a jejich funkcí.....	40
Tabulka 4: Porovnání možností nastavení systémů knihovnou .....	41
Tabulka 5: Porovnání možností nastavení systémů uživatelem.....	42



## Seznam zkratek

EAD	Encoded Archival Description
EDS	EBSCO Discovery Service
FTP	File Transfer Protocol
HILCC	Hierarchical Interface to Library of Congress Classification
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
MARC	MAchine-Readable Cataloging
MODS	Metadata Object Description Schema
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
OPAC	Online Public Access Catalog
RSS	Rich Site Summary
SFX	Special Effects
URL	Uniform Resource Locator
XML	Extensible Markup Language